Practically Useful Models for Kinetics of Biodiesel Production

Pulkit Chhabra,[†] Sebastian Mosbach,[‡] Iftekhar A. Karimi,[*,†] and Markus Kraft[*,‡,¶]

[†]*Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585, Singapore*

[‡]*Department of Chemical and Biotechnology, University of Cambridge, West Site, Philippa Fawcet Drive, Cambridge CB3 0AS, United Kingdom*

[¶]*School of Chemical and Biomedical Engineering, Nanyang Technological University, 62 Nanyang Drive, Singapore 637459, Singapore*

E-mail: cheiak@nus.edu.sg; mk306@cam.ac.uk

Abstract

We developed four kinetic models of varying complexity for biodiesel production. The models incorporate both transesterification and saponification, thereby making them practically applicable. We then propose an iterative parameter estimation algorithm to identify a prefixed number of significant rate constants via sensitivity analysis and estimate their kinetic parameters ($\mathbf{A}$ and $\Delta\mathbf{E}$) using non-linear regression. Using experimental data on eight different oils, two alcohols, and two catalysts, we show that our models accurately predict the dynamic concentration profiles of various species during the transesterification of oil. Furthermore, we demonstrate the applicability of the best model (based on the values of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Akaike Information Criterion (AIC)) for eleven additional experiments by predicting the final biodiesel properties with significant accuracy. Finally,

using N-way ANOVA, we identify the choice of oil, alcohol and catalyst as the most significant input factors followed by the operating conditions of the reactor.

Keywords

Transesterification, Saponification, Parameter Estimation, Sensitivity Analysis, ANOVA.

Introduction

The depletion of fossil fuel resources and increasing environmental concerns are driving the search for alternative fuels such as biodiesel that has gained enormous impetus in the past two decades. The earliest efforts for such alternative fuels can be traced to Rudolf Diesel a century ago. He was the first to test peanut oil as an alternative fuel for diesel engine.[1] During the 1930s and 1940s, vegetable oils were often used as alternative fuels in case of emergency.[2] However, researchers soon realized that they cannot directly substitute the petroleum-based diesel due to their high viscosities, low volatilities, and poor cold flow properties. Hence, they are usually processed to obtain biodiesels with properties similar to the petroleum-based diesel.[3]

Biodiesel is a non-toxic, renewable, and biodegradable promising alternative fuel composed mainly of fatty acid esters. Its global demand has increased ten-fold (IEA, 2014) over the last decade with annual production at 30 million m^3 in 2014.[4] It can be produced by at least four different processes: pyrolysis, dilution, micro-emulsion, and transesterification. Among these, transesterification is the most commonly adopted industrial process due to its higher conversion rates compared to the other processes.[5] Transesterification (also referred to as alcoholysis) is a process in which a triglyceride reacts with an alcohol in the presence of a catalyst to produce an alkyl ester and glycerol. The alkali catalysts such as sodium/potassium hydroxides are the most commonly used since the process is faster and process conditions are moderate. Although biodiesel is a promising alternative fuel, its

production poses significant challenges due to its higher production costs compared to the petroleum-based diesel. Therefore, it becomes crucial to model and optimize the biodiesel production process to lower costs while meeting the desired property specifications.

The literature is replete with research articles and reviews[6–8] on the modeling,[9–11] simulation,[12–14] process integration,[15–17] and optimization[11,17–20] of biodiesel production. The transesterification kinetics has been their primary goal with most studies focusing on the methanolysis[21–23] and ethanolysis[24–26] of specific common oils such as soybean,[27,28] sunflower,[29,30] and palm oils.[31,32] Their approach has been to develop lumped kinetic models that involve representing glycerides as one generic monoglyceride ($\overline{\text{MG}}$), one generic diglyceride ($\overline{\text{DG}}$), and one generic triglyceride ($\overline{\text{TG}}$). In other words, all the different types of a given glyceride are lumped into one generic glyceride ($\overline{\text{MG}}$, $\overline{\text{DG}}$, or $\overline{\text{TG}}$). Most conclude the reactions to be second order in both forward and backward directions. Noureddini and Zhu[27] and Bashiri and Pourbeiram[28] studied soybean oil transesterification with methanol in the presence of sodium hydroxide. They both proposed a 3-region kinetic model with an initial mass transfer controlled region followed by kinetics controlled reactions, and then equilibrium. Noureddini and Zhu[27] also examined the effects of mixing intensity and temperature. Vicente *et al*[30] studied the transesterification kinetics of sunflower oil. They reported that the mass-transfer controlled region is not significant at high impeller speed. Stamenovic *et al.*[29] reported a three-region model for sunflower oil. Darnoko and Cheryan[31] proposed a pseudo second order model in the initial stages followed by first order or zero order for the transetsrification of palm oil in the presence of potassium hydroxide. Shahbazi *et al.*[32] examined the transesterification kinetics of palm oil in the presence of both sodium and potassium hydroxides. They obtained a higher yield for the latter.

While the above models use $\overline{\text{MG}}$, $\overline{\text{DG}}$, and $\overline{\text{TG}}$, the kinetic parameters for the same reactions vary drastically from one study to another. For instance, the rate constants vary from 0.039 $\text{mol}^{-2}\text{L}^2\text{min}^{-1}$ to 9.09 $\text{mol}^{-2}\text{L}^2\text{min}^{-1}$ for $\overline{\text{TG}} \rightarrow \overline{\text{DG}}$; 0.012 $\text{mol}^{-2}\text{L}^2\text{min}^{-1}$ to 367 $\text{mol}^{-2}\text{L}^2\text{min}^{-1}$ for $\overline{\text{DG}} \rightarrow \overline{\text{MG}}$, and 0.0015 $\text{mol}^{-2}\text{L}^2\text{min}^{-1}$ to 11900 $\text{mol}^{-2}\text{L}^2\text{min}^{-1}$ for

$\overline{\text{MG}} \rightarrow \overline{\text{G}}$. This is most likely due to the fact that each study uses the experimental data on one specific oil as shown by Issrayikul et al.[33] who found the parameters for the same reaction to vary significantly between palm oil and mustard oil. This clearly suggests that kinetic parameters do vary with the type of glyceride and lumping different glycerides into one generic glyceride may not be accurate. Likozar and Levec[34,35] addressed this limitation by expressing oils in terms of all their different fatty acids. Using the experimental data on four oils, they predicted the dynamic concentration profiles of various glycerides and esters of fatty acids. They also studied the effects of oil, alcohol, catalyst, catalyst quantity, temperature, and alcohol-to-oil molar ratio on the biodiesel yields and oil conversions. However, their model is naturally huge and complex with many kinetic parameters whose accurate estimation remains a significant challenge. Furthermore, it is also not clear whether such a model complexity is warranted.

It is well known that saponification is a major unavoidable side reaction in biodiesel production. It can significantly affect the yield and performance of a biodiesel reactor. However, most existing studies ignore the presence of the saponification reactions. For instance, Likozar and Levec neglected the impact of saponification in their complex model, which limits its practical utility.

The above discussion leads us to the below questions on a practically useful kinetic model for biodiesel production.

1. What reactions should the model include?

2. Is it possible to have an accurate model that is not oil-specific?

3. Will the model vary with the catalyst (NaOH vs KOH)?

4. What species granularity is acceptable? In other words, should each glyceride and ester be considered as individually distinct or lumped species? How complex should the model be?

5. How can we estimate the potentially many kinetic parameters accurately and reliably?

In this work, we answer the above questions by performing the following tasks:

1. Develop a generalized reaction mechanism for biodiesel production from an arbitrary mix of glycerides, water, free fatty acid (FFA), alcohol (AOH), and base catalyst. The mechanism includes saponification as well.

2. Develop kinetic models of varying complexity for biodiesel production based on the above mechanism.

3. Devise an algorithm to identify and estimate the most important parameters in a complex kinetic model based on the available experimental data.

4. Validate the best model with the experimental data for specific oils.

Overall, we develop four generalized and practically useful models of varying complexity for the base-catalyzed kinetics of biodiesel production. Table S1 (Supporting Information) provides a comprehensive account of the factors considered in this work along with a comparison with two other studies in the literature.

This article is structured as follows. We first describe our kinetic models. We then explain our parameter estimation framework. This is followed by details regarding initial parameter estimates along with numerical analysis and comparison of our four models. Finally, we analyze the impact of key factors and their interactions on biodiesel yield and quality. In closing, we draw our conclusions and provide a brief perspective on future work.

Kinetic Models

While biodiesel can be produced from a variety of vegetable oils, most oils contain 95-98% of triglycerides (TG) with traces of diglycerides (DG), monoglycerides (MG), water, and free fatty acids (FFA). Therefore, without any loss of generality, we view a vegetable oil as a known mixture of these five components. A glyceride is simply one or more fatty acid groups attached to a glycerine (G) backbone. The glycerides from five fatty acid groups,

namely O = Oleic, L = Linoleic, N = Linolenic, P = Palmitic, and S = Stearic are the most common in oils. Thus, in this work, we consider the glycerides and esters of O, L, N, P, and S only. Now, diglycerides and triglycerides are called simple or mixed depending on whether their fatty acid groups are identical or different. For instance, triolein is a simple triglyceride with three oleic acid groups, which we label as OOOG. Distearin (SSG) is a simple diglyceride with two S groups. Dioleo-palmitin (OOPG) is a mixed triglyceride with two O groups and one P group. Thus, we consider five simple triglycerides (OOOG, LLLG, NNNG, PPPG, and SSSG), five simple diglycerides (OOG, LLG, NNG, PPG, and SSG), 10 mixed diglycerides (Table S2 in Supporting Information), 30 mixed triglycerides (Table S2), and five monoglycerides (OG, LG, NG, PG, and SG) as possible components in an oil in addition to water and FFA. The transesterification of these glycerides will then result in a biodiesel with 5 alkyl esters ($\mathbb{AE}$ = AOE, ALE, ANE, APE, ASE, where A is either methyl or ethyl) as possible components. We now use these species to propose four generalized models (M1, M2, M3, and M4) based on the degrees of lumping and complexity.

Model M1

M1 assumes that the rates of transesterification and saponification do not depend on the types of glycerides, esters, and FFAs. This allows us to represent all glycerides in terms of three generic glycerides ($\overline{\text{MG}}$, $\overline{\text{DG}}$, and $\overline{\text{TG}}$), all esters in terms of two generic esters ($\overline{\text{AE}}$, where A is methyl or ethyl), all soaps in terms of one generic soap ($\overline{\text{S}}$), and all FFAs in terms of one generic fatty acid ($\overline{\text{FFA}}$). Thus, M1 has eleven species, *viz.* ($\overline{\text{MG}}$, $\overline{\text{DG}}$, $\overline{\text{TG}}$, $\overline{\text{AE}}$, $\overline{\text{FFA}}$, $\overline{\text{S}}$, alcohol (AOH, methanol or ethanol), G, alkoxide (AO^-, methoxide or ethoxide), hydroxide (OH^-) and water (W), resulting in the following eight reactions.

$$\overline{\text{TG}} + \text{AOH} + \text{AO}^- \rightleftharpoons \overline{\text{DG}} + \overline{\text{E}} + \text{AO}^- \tag{R1.1}$$

$$\overline{\text{DG}} + \text{AOH} + \text{AO}^- \rightleftharpoons \overline{\text{MG}} + \overline{\text{E}} + \text{AO}^- \tag{R1.2}$$

$$\overline{\mathrm{MG}} + \mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{G} + \overline{\mathrm{E}} + \mathrm{AO^-} \tag{R1.3}$$

$$\mathrm{AOH} + \mathrm{OH^-} \rightleftharpoons \mathrm{AO^-} + \mathrm{W} \tag{R1.4}$$

$$\overline{\mathrm{TG}} + 3\mathrm{OH^-} \rightarrow 3\overline{\mathrm{S}} + \mathrm{G} \tag{R1.5}$$

$$\overline{\mathrm{E}} + \mathrm{OH^-} \rightarrow \overline{\mathrm{S}} + \mathrm{AOH} \tag{R1.6}$$

$$\overline{\mathrm{FFA}} + \mathrm{OH^-} \rightarrow \overline{\mathrm{S}} + \mathrm{W} \tag{R1.7}$$

$$\overline{\mathrm{FFA}} + \mathrm{AO^-} \rightarrow \overline{\mathrm{S}} + \mathrm{AOH} \tag{R1.8}$$

Here, the concentration of a generic species is obtained by summing the concentrations of all its constituent species. The above involve four reversible and four irreversible reactions, thereby leading to 24 kinetic parameters (12 activation energies and 12 pre-exponential factors).

Model M2

In contrast to M1, M2 simplifies the transesterification reactions. It assumes that reactions (R1.2) and (R1.3) are much faster compared to (R1.1), so a triglyceride converts directly into its corresponding esters during transesterification. However, M2 allows the rates to vary with the type of triglyceride. Thus, instead of lumping all triglycerides into one $\overline{\mathrm{TG}}$, it lumps them into five simple triglycerides (OOOG, LLLG, NNNG, PPPG, and SSSG), and considers five distinct FFAs (HO, HL, HN, HP, and HS). M2 has 10 reactions (6 reversible and 4 irreversible) involving 21 species as follows:

$$\mathrm{OOOG} + 3\mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{G} + 3\mathrm{AOE} + \mathrm{AO^-} \tag{R2.1}$$

$$\mathrm{LLLG} + 3\mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{G} + 3\mathrm{ALE} + \mathrm{AO^-} \tag{R2.2}$$

$$\mathrm{NNNG} + 3\mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{G} + 3\mathrm{ANE} + \mathrm{AO^-} \tag{R2.3}$$

$$\mathrm{PPPG} + 3\mathrm{AOH} + \mathrm{AO}^- \rightleftharpoons \mathrm{G} + 3\mathrm{APE} + \mathrm{AO}^- \tag{R2.4}$$

$$\mathrm{SSSG} + 3\mathrm{AOH} + \mathrm{AO}^- \rightleftharpoons \mathrm{G} + 3\mathrm{ASE} + \mathrm{AO}^- \tag{R2.5}$$

$$\mathrm{AOH} + \mathrm{OH}^- \rightleftharpoons \mathrm{AO}^- + \mathrm{W} \tag{R2.6}$$

$$\mathrm{OOOG} + \mathrm{LLLG} + \mathrm{NNNG} + \mathrm{PPPG} + \mathrm{SSSG} + 15\mathrm{OH}^- \rightarrow 15\overline{\mathrm{S}} + 5\mathrm{G} \tag{R2.7}$$

$$\mathrm{AOE} + \mathrm{ALE} + \mathrm{ANE} + \mathrm{APE} + \mathrm{ASE} + 5\mathrm{OH}^- \rightarrow 5\overline{\mathrm{S}} + 5\mathrm{AOH} \tag{R2.8}$$

$$\mathrm{HO} + \mathrm{HL} + \mathrm{HN} + \mathrm{HP} + \mathrm{HS} + 5\mathrm{OH}^- \rightarrow 5\overline{\mathrm{S}} + 5\mathrm{W} \tag{R2.9}$$

$$\mathrm{HO} + \mathrm{HL} + \mathrm{HN} + \mathrm{HP} + \mathrm{HS} + 5\mathrm{AO}^- \rightarrow 5\overline{\mathrm{S}} + 5\mathrm{AOH} \tag{R2.10}$$

where, AOE, ALE, ANE, APE, and ASE are 10 alkyl esters with A = methyl or ethyl. These 10 reactions involve 32 kinetic parameters (16 activation energies and 16 pre-exponential factors).

Model M3

M3 follows M2 in using the five simple triglyceride (OOOG, LLLG, NNNG, PPPG, and SSSG), but also uses the three stepwise transesterification reactions of M1. The five simple triglycerides then lead to five simple diglycerides (OOG, LLG, NNG, PPG, and SSG) and the five monoglycerides (OG, LG, NG, PG, SG) and five FFAs (HO, HL, HN, HP, and HS). Thus, M3 has 20 reactions (16 reversible and 4 irreversible) with 31 species as follows:

$$\mathrm{OOOG} + \mathrm{AOH} + \mathrm{AO}^- \rightleftharpoons \mathrm{OOG} + \mathrm{AOE} + \mathrm{AO}^- \tag{R3.1}$$

$$\mathrm{LLLG} + \mathrm{AOH} + \mathrm{AO}^- \rightleftharpoons \mathrm{LLG} + \mathrm{ALE} + \mathrm{AO}^- \tag{R3.2}$$

$$\mathrm{NNNG} + \mathrm{AOH} + \mathrm{AO}^- \rightleftharpoons \mathrm{NNG} + \mathrm{ANE} + \mathrm{AO}^- \tag{R3.3}$$

$$\mathrm{PPPG} + \mathrm{AOH} + \mathrm{AO}^- \rightleftharpoons \mathrm{PPG} + \mathrm{APE} + \mathrm{AO}^- \tag{R3.4}$$

$$\mathrm{SSSG} + \mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{SSG} + \mathrm{ASE} + \mathrm{AO^-} \tag{R3.5}$$

$$\mathrm{OOG} + \mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{OG} + \mathrm{AOE} + \mathrm{AO^-} \tag{R3.6}$$

$$\mathrm{LLG} + \mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{LG} + \mathrm{ALE} + \mathrm{AO^-} \tag{R3.7}$$

$$\mathrm{NNG} + \mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{NG} + \mathrm{ANE} + \mathrm{AO^-} \tag{R3.8}$$

$$\mathrm{PPG} + \mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{PG} + \mathrm{APE} + \mathrm{AO^-} \tag{R3.9}$$

$$\mathrm{SSG} + \mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{SG} + \mathrm{ASE} + \mathrm{AO^-} \tag{R3.10}$$

$$\mathrm{OG} + \mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{G} + \mathrm{AOE} + \mathrm{AO^-} \tag{R3.11}$$

$$\mathrm{LG} + \mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{G} + \mathrm{ALE} + \mathrm{AO^-} \tag{R3.12}$$

$$\mathrm{NG} + \mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{G} + \mathrm{ANE} + \mathrm{AO^-} \tag{R3.13}$$

$$\mathrm{PG} + \mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{G} + \mathrm{APE} + \mathrm{AO^-} \tag{R3.14}$$

$$\mathrm{SG} + \mathrm{AOH} + \mathrm{AO^-} \rightleftharpoons \mathrm{G} + \mathrm{ASE} + \mathrm{AO^-} \tag{R3.15}$$

$$\mathrm{AOH} + \mathrm{OH^-} \rightleftharpoons \mathrm{AO^-} + \mathrm{W} \tag{R3.16}$$

$$\mathrm{OOOG} + \mathrm{LLLG} + \mathrm{NNNG} + \mathrm{PPPG} + \mathrm{SSSG} + 15\mathrm{OH^-} \rightarrow 15\overline{\mathrm{S}} + 5\mathrm{G} \tag{R3.17}$$

$$\mathrm{AOE} + \mathrm{ALE} + \mathrm{ANE} + \mathrm{APE} + \mathrm{ASE} + 5\mathrm{OH^-} \rightarrow 5\overline{\mathrm{S}} + 5\mathrm{AOH} \tag{R3.18}$$

$$\mathrm{HO} + \mathrm{HL} + \mathrm{HN} + \mathrm{HP} + \mathrm{HS} + 5\mathrm{OH^-} \rightarrow 5\overline{\mathrm{S}} + 5\mathrm{W} \tag{R3.19}$$

$$\mathrm{HO} + \mathrm{HL} + \mathrm{HN} + \mathrm{HP} + \mathrm{HS} + 5\mathrm{AO^-} \rightarrow 5\overline{\mathrm{S}} + 5\mathrm{AOH} \tag{R3.20}$$

The above involve 72 kinetic parameters (36 activation energies and 36 pre-exponential factors).

Model M4

M4 is the most general of our four models, as it lumps only the soaps, and keeps all glycerides and esters distinct. Thus, M4 has 71 species (35 triglycerides (Table S2), 15 diglycerides (Table S2), 5 monoglycerides, 10 alkyl esters, 5 FFAs, $\overline{\mathrm{S}}$, AOH, G, AO^-, OH^- and W. It involves the 110 reactions (106 reversible and 4 irreversible, as shown in Table S3, provided in Supporting Information) with 432 kinetic parameters (216 pre-exponential factors and 216 activation energies).

The molar concentration (mol/L) profiles of various species in a constant-volume continuous-stirred batch reactor (CSBR) can be described by the following ODEs,

$$\frac{\mathrm{d}\mathbf{c}(t)}{\mathrm{d}t} = \mathcal{S}^T \cdot \mathbf{r}(\mathbf{k}, \mathbf{c}(t)) \qquad \mathbf{c}(t=0) = \mathbf{c}_0 \tag{E1}$$

$$\mathbf{k} = \mathbf{A} \exp\left(\frac{-\Delta\mathbf{E}}{\mathrm{RT}}\right) \tag{E2}$$

where $\mathbf{c}$(t) is the vector (11×1 for M1, 21×1 for M2, 31×1 for M3, and 71×1 for M4) of species concentrations, $\mathcal{S}$ is the stoichiometric coefficient matrix (12×11 for M1, 16×21 for M2, 36×31 for M3, and 216×71 for M3) for reactions (R1.1) to R4.110, and $\mathbf{r}(\mathbf{k}, \mathbf{c}(t))$ is the vector (12×1 for M1, 16×1 for M2, 36×1 for M3, and 216×1 for M4) of corresponding reaction rates (mol/L/min). Based on the findings reported in literature,[36] we assume that order is one with respect to all the species respectively.

Given $\mathbf{A}$, $\Delta\mathbf{E}$, $\mathbf{c}_0$, and a CSBR temperature profile T(t, $0 < t < t_f$), Eq. (E1) can be integrated numerically to obtain $\mathbf{c}(t)$ for $0 < t < t_f$. In this work, we used ODE15s in MATLAB R2015a to solve Eqs. (E1) and (E2). From $\mathbf{c}(t)$, the ester yield (EY) and triglyceride conversion (TGC) at time t can be computed as follows:

$$\text{Ester Yield } (\%) = \frac{\overline{\mathbf{E}}_t}{\left(3 \times \overline{\mathbf{TG}}_t\right) + \left(2 \times \overline{\mathbf{DG}}_t\right) + \left(1 \times \overline{\mathbf{MG}}_t\right)} \tag{E3}$$

$$\text{Triglyceride Conversion } (\%) = \left(1 - \frac{\overline{\mathbf{TG}}_t}{\overline{\mathbf{TG}}_0}\right) \times 100 \tag{E4}$$

where $\overline{\mathbf{E}}_t$ is the total concentration (mol/L) of esters and $\overline{\mathbf{TG}}_t$ is the total concentration (mol/L) of triglycerides at t.

Algorithm for Estimating Model Parameters

Accurate and plentiful experimental data are essential for reliably estimating the various model parameters. We reviewed the existing literature on biodiesel production to select 60 experiments (e=1, 2,...,60) that were suitable for our models. All were isothermal reactions, and involved eight oils (Cynara Cardunculus L.,[37] a 50:50 wt% blend of Jatropha oil and waste food oil,[38] Jatropha,[39] Linseed,[39] Palm,[33] Rapeseed,[36,40] Soybean,[41] and Sunflower[30,42]), two alcohols (Methanol and Ethanol), and two catalysts (NaOH and KOH). We classify these experiments into three data-sets: $\mathbb{DS}1$ with 18 experiments (e = 1,...,18) using methanol and KOH, $\mathbb{DS}2$ with 21 experiments (e = 19,...,39) using methanol and NaOH, and $\mathbb{DS}3$ with 21 experiments (e = 40,...,60) using ethanol and NaOH. The experiments with ethanol and KOH were too few and not suitable for our models. Each experiment e reports J_e reactor performances (Y_{ej}^{exp} = EY or TGC) at time points (j=1,2,...,J_e). Tables S4, S5, and S6 (Supporting Information) gives all the experimental conditions for each data-set $\mathbb{DS}1$, $\mathbb{DS}2$, and $\mathbb{DS}3$. In total, we have 181 Y_{ej}^{exp} values for $\mathbb{DS}1$, 169 for $\mathbb{DS}2$, and 153 for $\mathbb{DS}3$.

Given the limited experimental data, it is not prudent to estimate all the model parameters simultaneously, as many are expected to be inconsequential. Therefore, we need to identify the most significant parameters for each model, and then estimate their best-fit values. To this end, we propose an iterative parameter estimation algorithm shown in Figure 1. The algorithm assumes a prefixed number (NP) for the significant rate constants, and begins with some initial estimates for all the model parameters. Each iteration involves two steps. In the first step, we identify NP significant rate constants based on a sensitivity analysis. In

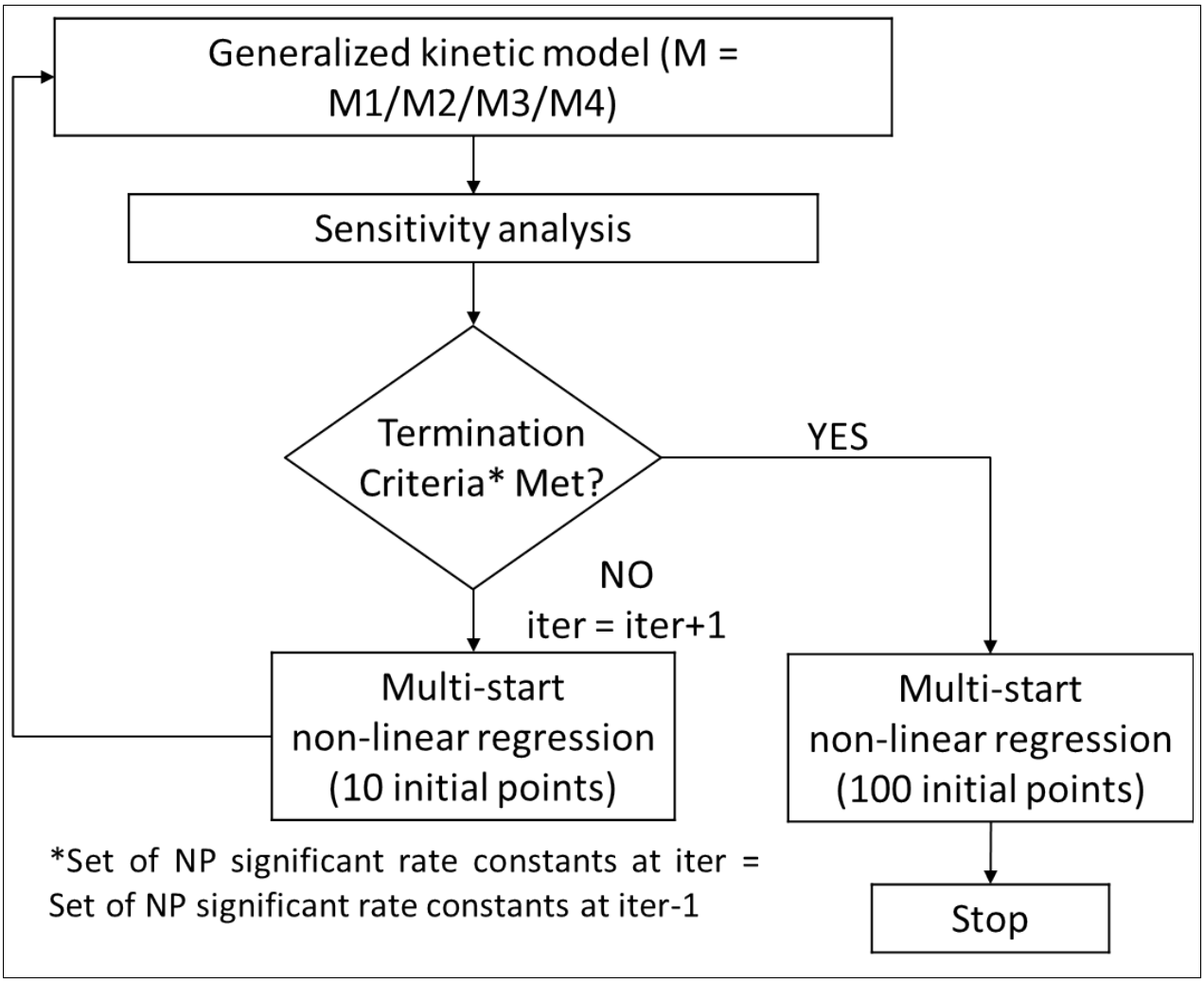


Figure 1: Adopted Framework for parameter estimation.

the second step, we estimate the best values for the activation energies and pre-exponential factors corresponding to those NP rate constants by using nonlinear regression. Since the sensitivities of the rate constants may depend on the assumed parameter values, and the identities of the significant rate constants are not known with surety, we need to perform iterations until these two are consistent with each other. The iterations will allow the set of significant rate constants to change in response to the best-fit parameter values over iterations. We terminate our iterations, when the significant set remains unchanged in two successive iterations.

Sensitivity analysis

Consider estimating the parameters ($\mathbf{A}$, $\Delta\mathbf{E}$, and $\mathbf{k}$) of a model M = M1, M2, M3, or M4 from a given data set DS = $\mathbb{DS}1$, $\mathbb{DS}2$, or $\mathbb{DS}3$. Let $\mathbf{A}^*$, $\Delta\mathbf{E}^*$, and $\mathbf{k}^*$ denote the best estimates for $\mathbf{A}$, $\Delta\mathbf{E}$, and $\mathbf{k}$ at an iteration in our algorithm. Define an average normalized sensitivity coefficient **NSC** for a rate constant k as follows:

$$\begin{aligned}\mathbf{NSC}(\mathrm{k}) = \frac{1}{J}\sum_{e\in DS}\sum_{j=1}^{J_e}\left(\frac{Y_{ej}\left(\mathrm{k}=\mathrm{k_{ub}}\right)-Y_{ej}\left(\mathrm{k}=\mathrm{k_{lb}}\right)}{|\mathrm{k_{ub}}-\mathrm{k_{lb}}|}\right)\\ \times\left(\frac{\sqrt{\mathrm{k_{ub}}\times\mathrm{k_{lb}}}}{\sqrt{Y_{ej}\left(\mathrm{k}=\mathrm{k_{ub}}\right)\times Y_{ej}\left(\mathrm{k}=\mathrm{k_{lb}}\right)}}\right)\end{aligned} \tag{E5}$$

where, $J = 181$ for $\mathbb{DS}1$, $J = 169$ for $\mathbb{DS}2$, and $J = 153$ for $\mathbb{DS}3$; $\mathbf{k}_{lb} = \mathbf{A}_{lb}\exp\left(\frac{-\Delta\mathbf{E}_{ub}}{\mathrm{RT}}\right)$ refers to the lower bound on the rate constant; and $\mathbf{k}_{ub} = \mathbf{A}_{ub}\exp\left(\frac{-\Delta\mathbf{E}_{lb}}{\mathrm{RT}}\right)$ refers to the upper bound on the rate constant. We rank the rate constants in the decreasing order of their NSC values, and identify the first NP rate constants as the most significant for the current iteration.

Nonlinear Regression

Let us consider a model M = M1, M2, M3, or M4. We set the $\mathbf{A}$ and $\Delta\mathbf{E}$ of all rate constants at their current best estimates ($\mathbf{A}^*$, $\Delta\mathbf{E}^*$, and $\mathbf{k}^*$). For a particular iteration, we keep all the parameters fixed except the ones corresponding to the NP significant rate constants. We then use fmincon function in MATLAB to minimize the sum of squared errors SS = SS1, SS2, and SS3 separately by varying these $\mathbf{A}(\mathrm{k})$ and $\Delta\mathbf{E}(\mathrm{k})$ between $0.1\mathbf{A}^*(\mathrm{k}) \leq \mathbf{A}^*(\mathrm{k}) \leq 10\mathbf{A}^*(\mathrm{k})$ and $0.8\Delta\mathbf{E}^*(\mathrm{k}) \leq \Delta\mathbf{E}^*(\mathrm{k}) \leq 1.2\Delta\mathbf{E}^*(\mathrm{k})$, where SS1 is the sum of squared errors for $\mathbb{DS}1$, and likewise.

Given the highly nonlinear nature of the above optimization problem, *fmincon* cannot guarantee a globally best solution. Hence, we use a multi-start strategy to maximize our chances of getting such a solution. We solve the nonlinear regression problem with 10

different initial estimates for the 2NP parameters ($\mathbf{A}$ and $\Delta\mathbf{E}$) at each iteration except for the last, where we use 100 initial estimates. We generate the initial estimates from the Sobol sequence in MATLAB. The solution with the least SS gives us the best parameter estimates for the next iteration.

We implemented the algorithm in MATLAB R2015a on a Dell Precision T-7910 with Intel(R) Xeon(R) processor (36 cores, 2.3 GHz), 64.0 GB RAM and 64-bit Windows-7 OS.

Model Parameters and Comparison

We define two performance measures (Root Mean Square Error (RMSE) and Mean Absolute Error (MAE)) to assess each model for each data set ($\mathbb{DS}1$, $\mathbb{DS}2$, or $\mathbb{DS}3$).

$$RMSE = \sqrt{\frac{1}{J}\sum_{j=1}^{J}\left(Y_{ej}^{exp} - Y_{ej}\right)^2} \tag{E6}$$

$$MAE = \frac{1}{J}\sum_{j=1}^{J}\left(\left|Y_{ej}^{exp} - Y_{ej}\right|\right) \tag{E7}$$

Parameters for M1

We used the values of $\mathbf{A}$ and $\Delta\mathbf{E}$ from Valentine *et. al.*[36] as our initial estimates. The algorithm took 4 iterations for $\mathbb{DS}1$, 5 iterations for $\mathbb{DS}2$, and 4 iterations for $\mathbb{DS}3$. Tables S8-S10 (Supporting Information) show how the significant rate constants and their best-fit parameter values change with iterations for each data-set. M1 has MAE and RMSE = 5.8% and 7.6% for $\mathbb{DS}1$, 5.1% and 6.4% for $\mathbb{DS}2$, and 5.4% and 7.3% for $\mathbb{DS}3$ (see Table 1).

Parameters for M2

To select the initial parameter estimates for our algorithm, we first identify the transesterification reactions ((R1.1) to (R1.3)) in M1 with the smallest rate constant (Eq. (E2)) at

333.15 K. While 333.15 K seems to be the most common temperature in the experimental data, the slowest reaction remains the same for all practical temperatures, but varies with $\mathbb{DS}1$, $\mathbb{DS}2$, or $\mathbb{DS}3$. Since M2 assumes that reactions (R1.2) and (R1.3) are much faster than (R1.1), it makes sense to assign the smallest rate constant as the initial estimate. Since reactions (R2.1) to (R2.5) are similar, we can use the same initial estimates for them. Thus, the parameters for the smallest rate constant in M1 are used for reactions (R2.1) to (R2.5). Similarly, reactions (R1.4) to (R1.8) are the same as (R2.6) to (R2.10), so we use the best-fit parameter values of the former directly as the initial estimates for the latter. The algorithm took 2 iterations for $\mathbb{DS}1$, 8 iterations for $\mathbb{DS}2$, and 2 iterations for $\mathbb{DS}3$. Tables S11-S13 (Supporting Information) show how the significant rate constants and their best-fit parameter values change with iterations for each data-set. M2 has MAE and RMSE = 3.2% and 4.4% for $\mathbb{DS}1$, 3.3% and 4.3% for $\mathbb{DS}2$, and 6.1% and 7.9% for $\mathbb{DS}3$ (see Table 1). Clearly M2 is performing better for $\mathbb{DS}1$ and $\mathbb{DS}2$. While M1 is performing better than M2 for $\mathbb{DS}3$.

Parameters for M3

Since M3 has some similarities with M1 and M2, we utilize the previous information from M1 and M2 to select the initial parameter estimates for M3. For instance, the parameters for reactions (R2.1) to (R2.5) (TG to AE) can be used as the initial estimates for (R3.1) to (R3.5) (TG to DG). Since reactions (R3.6) to (R3.10) are specific versions of (R1.2), the parameters for (R1.2) are used. Similarly, the parameters of (R1.3) are used for (R3.11) to (R3.15) and those for (R2.6) to (R2.10) are used for (R3.16) to (R3.20). The algorithm took 4 iterations $\mathbb{DS}1$, 6 iterations for $\mathbb{DS}2$, and 5 iterations for $\mathbb{DS}3$. Tables S14-S16 (Supporting Information) show how the significant rate constants and their best-fit parameter values change with iterations for each data-set. M3 has MAE and RMSE = 2.5% and 3.7% for $\mathbb{DS}1$, 3.1% and 4.1% for $\mathbb{DS}2$, and 5.9% and 7.9% for $\mathbb{DS}3$ (see Table 1). M3 is performing the best for $\mathbb{DS}1$ and $\mathbb{DS}2$, but M1 is the best for $\mathbb{DS}3$.

Parameters for M4

Since M4 is a generalized version of M3, we can use the best-fit parameter values from M3 to select some initial estimates for M4. For these estimates, we can assume that the kinetics depends only on the type of fatty acid chain that reacts rather than its position in the glyceride molecule. Therefore, OLPG $\rightleftharpoons$ OPG and OLLG $\rightleftharpoons$ OLG will have the same initial parameter estimates as LLLG $\rightleftharpoons$ LLG. This enables us to use the parameters for reactions (R3.1) to (R3.15) as the initial estimates for R4.1 to R4.105 (Table S3 in Supporting Information), and those for (R3.16) to (R3.20) for R4.106 to R4.110 (Table S3). The algorithm took 7 iterations for $\mathbb{DS}1$, 10 iterations for $\mathbb{DS}2$, and 2 iterations for $\mathbb{DS}3$. Tables S17-S19 (Supporting Information) show how the significant rate constants and their best-fit parameter values change with iterations for each data-set. M4 has MAE and RMSE = 2.6% and 4.5% for $\mathbb{DS}1$, 5.1% and 6.4% for $\mathbb{DS}2$, and 5.7% and 8.0% for $\mathbb{DS}3$ (see Table 1). Clearly M4 is not the best model for any of the three data sets.

Some sources of prediction errors can be the following,

- The experimental studies give limited or no information on oil compositions. We had to assume a single uniform composition for each oil, which we assumed from one single source.[43] Clearly, the composition may vary from one experiment to another even for the same oil.

- We assumed mass transfer effects to be negligible. While this is a sound practical assumption and has been shown to have no/negligible effect at high mixing intensity, there is no way to guarantee that this was true for each and every experiment/oil.

- Most studies reported plots rather than tabular data. We had to use an online digitizer to obtain data at specific times.

- While we have tried our best to ensure global optimization, it cannot be guaranteed for this highly nonlinear and complex regression problem.

Model Comparison

While both MAE and RMSE measure average model prediction errors, RMSE is higher, when there is higher variance in frequency distribution of error magnitudes. In other words, higher RMSE may indicate the presence of outliers. Hence, as suggested by Garud *et al.*,[44] we combine these two measures to obtain a normalized Pooled Error ($\widehat{\text{PE}}_M$) to compare M = M1, M2, M3, and M4 for each data set (DS = $\mathbb{DS}1$, $\mathbb{DS}2$, and $\mathbb{DS}3$) as follows:

$$\widehat{\text{PE}}_M = \frac{\text{PE}_M}{\min\left(\text{PE}_{M1}, \text{PE}_{M2}, \text{PE}_{M3}, \text{PE}_{M4}\right)} \tag{E8}$$

$$\widehat{\text{PE}}_M = \sqrt{\text{MAE}_M \times \text{RMSE}_M} \tag{E9}$$

Additionally, we use the normalized Akaike Information Criterion ($\widehat{\text{AIC}}_M$) to evaluate our models. This is because the $\widehat{\text{PE}}_M$ values do not account for the model complexity and therefore may favor over-fitting. AIC is a well known metric that penalizes extra parameters while ranking the models in terms of the sum of squared errors. Mathematically this can be stated as follows,

$$\text{AIC}_M = 2n + Jln(SS) \tag{E10}$$

$$\widehat{\text{AIC}}_M = \frac{\text{AIC}_M}{\min\left(\text{AIC}_{M1}, \text{AIC}_{M2}, \text{AIC}_{M3}, \text{AIC}_{M4}\right)} \tag{E11}$$

where, n is the number of kinetic parameters for model M (n = 24 for M1, 32 for M2, 72 for M3, and 432 for M4), J is the number of data-points, and SS (= SS1, SS2, SS3) is the sum of squared errors for data-sets $\mathbb{DS}1$, $\mathbb{DS}2$, and $\mathbb{DS}3$ respectively. Finally, we combine these two to obtain an overall metric for each model M = M1, M2, M3, and M4 (OM_M) as follows,

$$OM_M = \widehat{\text{AIC}}_M \times \widehat{\text{PE}}_M \tag{E12}$$

As shown in Table 1, we see that M3 is the best model for $\mathbb{DS}1$ and $\mathbb{DS}2$. M1 seems too simple and M4 too complex for our purpose and the available data. In other words, simple

Table 1: Metrics indicating the performance of M1, M2, M3, and M4 for each data set $\mathbb{DS}1$, $\mathbb{DS}2$, and $\mathbb{DS}3$.

Data Set	Model	MAE	RMSE	$\widehat{\text{PE}}$	$\widehat{\text{AIC}}$	OM_M	Best Model
Set-1	M1	5.8	7.6	2.20	1.12	2.46	M3
	M2	3.2	4.4	1.27	1.00	1.27	
	M3	2.5	3.7	1.00	1.02	1.02	
	M4	2.6	4.5	1.13	1.53	1.73	
Set-2	M1	5.1	6.4	1.58	1.08	1.71	M3
	M2	3.3	4.3	1.06	1.00	1.06	
	M3	3.1	4.1	1.00	1.05	1.05	
	M4	5.1	6.4	1.58	1.65	2.61	
Set-3	M1	5.4	7.3	1.00	1.00	1.00	M1
	M2	6.1	7.9	1.10	1.03	1.13	
	M3	5.9	7.9	1.08	1.08	1.17	
	M4	5.7	8.0	1.08	1.59	1.72	

lumping of glycerides or esters into single generic ones as in M1 does not seem a good idea. Similarly, the other extreme of keeping them all distinct is also not warranted. Neglecting the diglycerides and monoglycerides might have deteriorated the performance for M2. However, this reasoning does not seem to extend to $\mathbb{DS}3$, as M1 is the best model for $\mathbb{DS}3$. This may be due to the lack of sufficient data for $\mathbb{DS}3$ as far as M2, M3, and M4 are concerned. If more detailed data were available, then the situation could have been different. However, performance of M3 is comparable to M1 even for $\mathbb{DS}3$, hence M3 can be regarded as the best model.

The superior performance of M3 versus M1 can be explained by the fact that M3 accounts for the finer composition of oil, while M1 does not. However, this logic does not extend to M4, as M3 performs better than M4. This can be due to one or more of the following. First, the experimental data used in this study involves EY or TGC at various times. Many different combinations of ester or triglyceride concentrations can produce the same result. As the number of glycerides increases, these combinations increase exponentially, making it difficult to obtain the globally best-fit model. Second, the experimental studies provide no information on their oil compositions or report only the fatty acid contents. While this information is sufficient for M2 and M3, it is not for M4. We had to assume oil compositions

from a single study,[43] which may not be accurate for all the experimental studies. Thus, the data limitations may have reduced the effectiveness of M4.

Let us now see how our models perform on some oil-alcohol-catalyst combinations that did not feature in our 60 experiments. We found 11 ($e = 1, 2, \ldots, 11$) such experiments (Table S7 in Supporting Information). These 11 experiments reported only the final biodiesel properties, and no detailed reaction profiles. Since M1 considers all glycerides and esters as generic species, it cannot provide any information on the various ester species in the final product. Hence, M1 cannot be tested for these 11 experiments. So, we use M2, M3, and M4 to predict the final biodiesel compositions. The literature correlations[45] in Table S20 (Supporting Information) are then used to estimate the biodiesel properties in Table S21 (Supporting Information). We observe that the models exhibit comparable performance (as shown in Table S21). However, since M2 and M3 are far less complex than M4, they are considered as better alternatives. Among M2 and M3, M3 is preferred because of its ability to predict the diglyceride and monoglyceride composition, which is a critical standard specification for biodiesel.

Figures S1-S3 (Supporting Information) show the predictions and errors from M3 versus experimental observations. Key observations from the plots are as follows:

- The triglyceride conversion increases monotonically with reaction temperature. Thus, the forward reactions dominate the reverse reactions even at higher temperatures. Furthermore, ester yield also increases with temperature. In other words, the saponification reactions are not decreasing the yield at higher temperatures and hence are not significant as far as temperature effects are concerned.

- The ester yields go through maxima as catalyst concentration increases. This can be explained as follows. The concentration of hydroxide ions increases with the catalyst concentration. For a fixed amount of alcohol, the excess hydroxide ions promote the saponification of esters and triglycerides and reversal of hydroxide-alkoxide reactions, which decrease ester yields.

- The ester yield or TG conversion increases monotonically with alcohol to oil ratio. Higher alcohol amounts lead to higher alkoxide ion concentrations, promoting transesterification and retarding saponification.

Based on our comprehensive evaluation, we ascertain that M3 is the best model in terms of prediction ability. We now analyze its robustness against variation in the triglyceride composition. This is because the triglyceride composition not only varies among oils but also across oils from different sources. Hence, it is vital that the model performance is not significantly affected due to such variation. To this end, we randomly generate samples for the triglyceride percentage within the range specified by Andrikopoulos (2002),[43] for each oil featuring in the 60 experiments (Table S22 in Supporting Information). We then evaluate the performance of M3 against different oil compositions by using PE. As shown in Table S23, we observe a maximum deviation of approximately 0.5% in the obtained results. Clearly, the model exhibits robust performance, thereby, reinforcing our confidence in the practical applicability of our model.

Identification of Key Factors and Interactions for Biodiesel Yield and Quality.

In the final part of our study, we use M3 to study the impact of various input factors (e.g. temperature (IF1), catalyst concentration (IF2), alcohol to oil molar ratio (IF3)) on Ester Yield (EY) and Pour Point (PP). Table S24 (Supporting Information) shows the nine factors and their levels. To this end, we conduct a full factorial design of experiments using M3 and perform an analysis of variance (ANOVA) using the *anovan* function in MATLAB. Our design involves 67,500 experiments on M3 for each of $\mathbb{DS}1$, $\mathbb{DS}2$ and $\mathbb{DS}3$. Table S25 (Supporting Information) lists the p-values and F-statistic for each factor and their interactions. Such an extensive analysis will be impractical with real experiments, but having a generalized model such as M3 makes it possible. Our analysis reveals the following,

- All nine factors have zero p-values, hence they are all significant. However, based on their F-statistics, we can conclude their relative order of significance as IF9 > IF8 > IF3 > IF1 > IF6 > IF2 > IF7 > IF5 > IF4 for yield, and IF8 > IF9 > IF7 > IF3 > IF1 > IF6 > IF2 > IF5 > IF4 for PP. We observe that the choice of oil (IF7) is relatively more important for PP than EY. This is because the biodiesel properties are governed by its composition which is strongly dependent on the choice of oil. We can say that all 2-factor interactions except those involving IF5 and IF6 are significant. Finally, the 3-factor interactions are mostly insignificant and hence can be neglected. The p-values associated with terms involving both IF8 and IF9 could not be computed due to missing set of experiments for ethanol-KOH (alcohol-catalyst) combination.

- The p-values and F-statistic techniques test whether a set of group means are equal or not. Rejection of the null hypothesis leads to the conclusion that not all group means are the same. This result, however, does not provide further information on which group means are different and how much different are they. To this end, we use the *multcompare* function in MATLAB to compute population marginal means of the two responses for each level of the nine input factors. The population marginal mean of a particular factor, say IF1 is computed by averaging out the effects of other factors i.e. IF2, ..., IF8. Although the best way to compute these values is to conduct a one-shot analysis for all the nine factors, the lack of data for ethanol-KOH combination enforces us to compute them in two parts *viz.* (1) computing population marginal means for IF1, ..., IF7, IF8 while setting IF9 = NaOH; (2) computing population marginal means for IF1, ..., IF7, IF9 while keeping IF8 = Methanol. Figures 2 show the population marginal means of the two responses versus the factorial design levels of the input factors. Key remarks from the analysis are as follows:

    1. The EY increased, while PP decreased with increase in IF1, IF2, IF3, and IF4 (Figures 2(a) and 2(c)). This is in accord with the experimental observations for

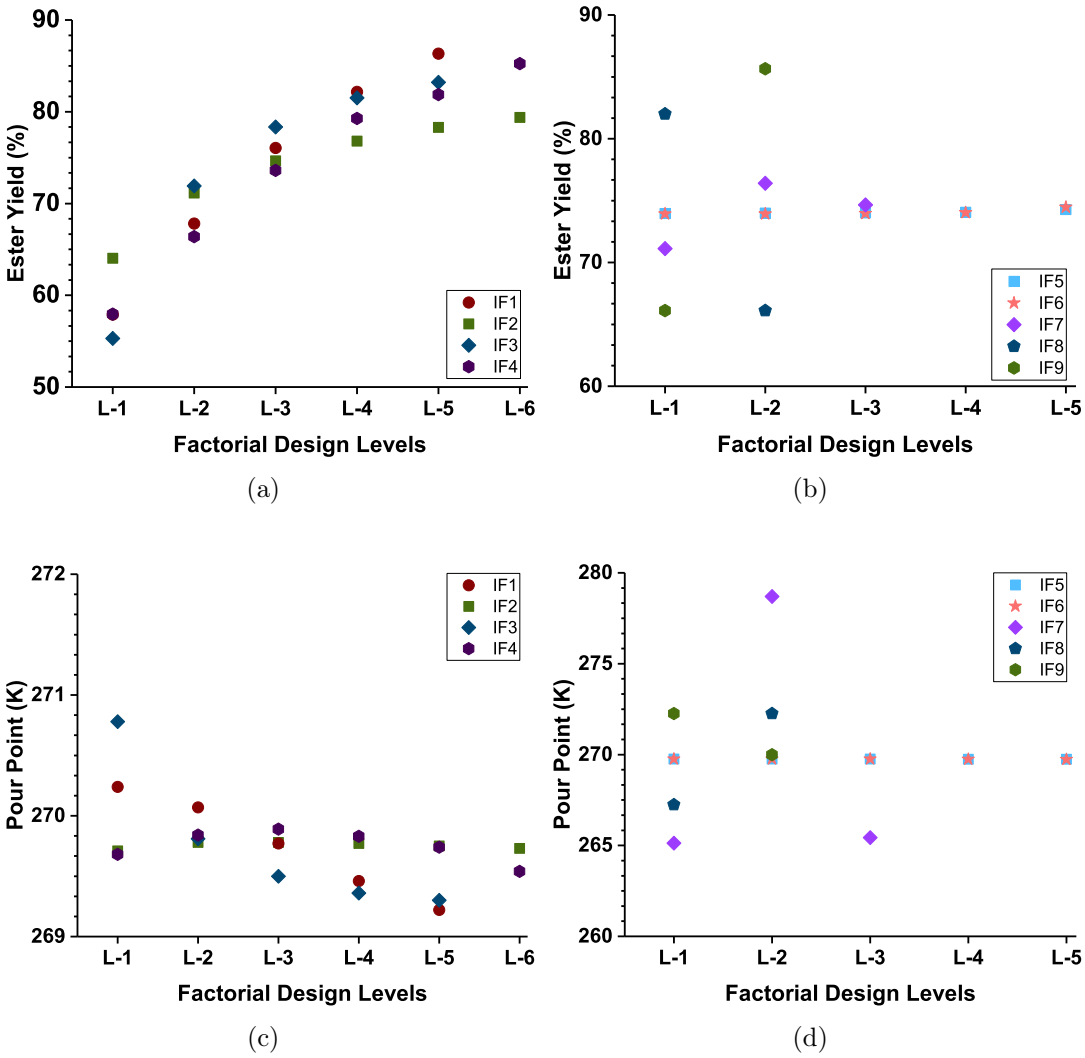


Figure 2: Effects of the nine input factors on the Ester Yield and Pour Point.

IF1, IF3, and IF4. However, unlike the experimental data, we did not observe any maxima for IF2.

2. The changes in the responses are negligible with respect to the changes in IF5 and IF6 (Figures 2(b) and 2(d)). Thus this supports the results indicated by the F-statistics.

3. The choice of oil (IF7) did not affect the EY and CN significantly. However, the PP values are much higher for Cottonseed oil than for the other two oils.

4. As stated earlier, the mean responses for IF8 and IF9 have been obtained in

two separate analysis. Thus, even though the plots suggest that higher yield is obtained using ethanol as alcohol and KOH as catalyst, it does not indicate that ethanol-KOH form the best alcohol-catalyst combination. Instead, the correct way to interpret is that if IF8 = Methanol, then IF9 = KOH can provide a higher yield. Similarly, when IF9 = NaOH then IF8 = Ethanol can provide a higher yield.

5. We also notice that varying the reactor operating conditions (IF1 to IF4) significantly affect the yield but do not have a remarkable impact on the biodiesel properties. On the other hand, the opposite is true for IF7. Hence, the significance of input factors strongly depends on the response. Therefore, given a set of inputs and outputs, only a subset of the input factors have a strong impact on a subset of the responses. Hence, selecting an appropriate subset is quite crucial. This is where a generalized model like the one developed in this work can have a huge impact.

Conclusion

We developed four practically useful kinetic models viz. M1, M2, M3, and M4 for the kinetics of biodiesel production. These models vary in complexity in terms of number of reactions that range from 8 reactions (4 reversible and 4 irreversible) for M1 to 110 reactions (106 reversible and 4 reversible) for M4, number of species that range from 11 for M1 to 71 for M4, and the number of kinetic parameters that range from 24 for M1 to 216 for M4. Thus, M1 is the simplest while M4 is the most complex model. We then propose an iterative algorithm to identify a prefixed number of significant rate constants and estimate their kinetic parameters ($\mathbf{A}$ and $\Delta\mathbf{E}$). The key novelties of this study are: (1) for a given set of data, we analyze the impact of both generality and complexity in the kinetics of biodiesel production, (2) our models incorporate both transesterification and saponification and are

therefore practical, (3) we propose an iterative parameter estimation algorithm for estimating the kinetic parameters of the four models, (4) we compare and analyze the performance of the four models for a variety of experimental settings, and (5) we perform comprehensive analysis for identification of key factors and their interactions using N-way Analysis of Variance.

Our extensive numerical analysis using dynamic data *viz.* ester yield or triglyceride conversion for eight oils, two alcohols, and two catalysts and static data i.e. biodiesel properties for eleven oils, two alcohols and two catalysts indicate that our models perform significantly well. We observe that overall M3 is the best performing model based on its applicability and multiple performance metrics *viz.* Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Pooled Error (PE), Akaike Information Criterion (AIC), and an Overall Metric (OM). Finally, we use our best model to identify the key factors and their interactions for Ester Yield and Pour Point. Our analysis indicates that if one aims to optimize the yield, the reactor operating conditions could be selected as the decision variables. On the other hand, if the performance metric is a function of the biodiesel properties, then selection of an oil or alcohol or catalyst may have a far greater impact. We also notice that higher yield could be obtained with methanol-KOH and ethanol-NaOH alcohol-catalyst combinations. Similar conclusion could not be provided for the choice of oil as the number of oils that are most commonly used throughout the world is far greater than either the alcohol or catalyst. Furthermore, considering the fact that the oils can be blended in varying proportions make it far more difficult to provide conclusive evidence with respect to the same. Therefore, in the future, we aim to use the generalized model developed in this study to analyze the impact of oils on both the yield and quality of biodiesel. Since feedstock contributes approximately 60% to 80% of the total production cost, we believe detailed understanding of the impact of an oil on the biodiesel production process can aid in making biodiesel a more economically viable candidate in the fuel market.

Acknowledgement

This project is funded by the National Research Foundation (NRF), Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

Supporting Information Available

- Filename: Practically Useful Models for Kinetics of Biodiesel Production - Supporting Information.
- Number of pages: 26 (25 tables and 3 figures).
- Description: This file includes the following information,
    1. A comprehensive account of the factors considered in this work (Table S1).
    2. List of triglyceride and diglyceride components in Model M4 (Table S2).
    3. List of transesterification and Saponification reactions for Model M4 (Table S3).
    4. Experimental details for the 60 experiments used for fitting the model (Tables S4-S6) and the 11 experiments used for validation (Table S7).
    5. List of significant rate constants along with their best fit parameter values at each iteration for each model M = M1, M2, M3, and M4 and each data-set DS = $\mathbb{DS}1$, $\mathbb{DS}2$, and $\mathbb{DS}3$ (Tables S8-S19).
    6. Plots depicting the fit of experimental data and Model predictions by M3 for each of the 60 experiments used for fitting the models along with performance measures (Figures S1-S3).
    7. Literature correlations for computing the biodiesel properties (Table S20).
    8. Experimental observations versus model predictions for eleven additional experiments (Table S21).

9. Triglyceride composition (Mean±Standard Deviation) for the eight oils featuring in the 60 experiments (Table S22).
10. Pooled Error (Mean±Standard Deviation) computed for evaluating the robustness of M3 against variations in the triglyceride composition reported in Table S22 (Table S23).
11. List of input Factors and their Factorial Design Levels (Table S24).
12. F-statistics and p-values of the nine input factors (IF1,. . . ,IF9) for Ester Yield and Pour Point computed using ANOVA (Table S25).

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

Notations and Symbols

Sets

$\mathbb{AE}$: Set of Alkyl Esters

$\mathbb{DS}1$: Set of experiments for Methanol-KOH combination.

$\mathbb{DS}2$: Set of experiments for Methanol-NaOH combination.

$\mathbb{DS}3$: Set of experiments for Ethanol-NaOH combination.

Indices

e: Experiment ($e = 1, \ldots, 60$)

j: Data-points (number of data-points vary with experiment number, see Tables S4, S5, and S6)

ae: Alkyl Ester (ae = AOE, ..., ASE, where A is either methyl or ethyl)

M: Model (M = M1, ..., M4)

Symbols

$\overline{\text{TG}}$: Generic Triglyceride

$\overline{\text{DG}}$: Generic Diglyceride

$\overline{\text{MG}}$: Generic Monoglyceride

$\overline{\text{FFA}}$: Generic Free Fatty Acid

$\overline{\text{AE}}$: Generic Alkyl Ester

$\overline{\text{S}}$: Generic Soap

G: Glycerine

W: Water

AOH: Alcohol (A is either methyl or ethyl)

AO^-: Alkoxide (A is either methyl or ethyl)

OH^-: Hydroxide

O: Oleic acid

L: Linoleic acid

N: Linolenic acid

P: Palmitic acid

S: Stearic acid

$\mathbf{c}(t)$: Species concentration vector (11$\times$ 1 for M1, 21$\times$ 1 for M2, 31$\times$ 1 for M3, and 71$\times$ 1 for M4)

$\mathbf{c}_0$: Initial Concentration vector

$\mathcal{S}$: Stoichiometric coefficient matrix (12$\times$11 for M1, 16$\times$21 for M2, 36$\times$31 for M3, and 216$\times$71 for M4)

$\mathbf{r}(\mathbf{k}, \mathbf{c}(t))$: Reaction rate vector (12$\times$ 1 for M1, 16$\times$ 1 for M2, 36$\times$ 1 for M3, and 216$\times$ 1 for M4)

$\mathbf{k}$: Vector of rate constants

k_{lb}: Lower bound for rate constant 'k' (computed at $0.1\mathbf{A}(\text{k})$ and $1.2\Delta\mathbf{E}(\text{k})$)

k_{ub}: Upper bound for rate constant 'k' (computed at $10\mathbf{A}(\text{k})$ and $0.8\Delta\mathbf{E}(\text{k})$)

A: Vector of pre-exponential factors

Δ**E**: Vector of activation energies

T: Reaction temperature (K)

R: Universal gas constant (R = 0.008314 kJ/mole K)

t: time-point

t_f: Final time-point

Y_{ej}: Model prediction (EY or TGC for experiment e at time-point j)

Y_{ej}^{exp}: Experimental response (EY or TGC for experiment e at time-point j)

NSC(k): Normalized sensitivity coefficient of rate constant k

NP: Number of most significant rate constants

J_e: Number of data-points for experiment 'e'

J: Total number of data-points (J = 181 for $\mathbb{DS}1$, 169 for $\mathbb{DS}2$, and 153 for $\mathbb{DS}3$)

MAE_M: Mean Absolute Error for model 'M'

RMSE_M: Root Mean Square Error for model 'M'

PE_M: Pooled Error for model 'M'

AIC_M: Akaike Information Criterion for model 'M'

$\widehat{\text{PE}}_M$: Normalized Pooled Error for model 'M'

n: Number of kinetic parameters (A and ΔE)

$\widehat{\text{AIC}}_M$: Normalized Akaike Information Criterion for model 'M'

OM_M: Overall Metric for model 'M'

NC_{ae}: Number of carbon atoms with respect to the fatty acid associated with alkyl ester 'ae'

PW_{ae}: Percentage weight of alkyl ester 'ae'

NDB_{ae}: Number of double bonds with respect to the fatty acid associated with alkyl ester 'ae'

ρ: Density

μ: Viscosity

References

(1) Shay, E. G. Diesel fuel from vegetable oils: status and opportunities. *Biomass and Bioenergy* **1993**, *4*, 227–242, DOI: 10.1016/0961-9534(93)90080-N.

(2) Singh, S.; Singh, D. Biodiesel production through the use of different sources and characterization of oils and their esters as the substitute of diesel: a review. *Renewable and Sustainable Energy Reviews* **2010**, *14*, 200–216, DOI: 10.1016/j.rser.2016.04.054.

(3) Verma, P.; Sharma, M. Review of process parameters for biodiesel production from different feedstocks. *Renewable and Sustainable Energy Reviews* **2016**, *62*, 1063–1071, DOI: 10.1016/j.rser.2009.07.017.

(4) Sajjadi, B.; Raman, A. A. A.; Arandiyan, H. A comprehensive review on properties of edible and non-edible vegetable oil-based biodiesel: Composition, specifications and prediction models. *Renewable and Sustainable Energy Reviews* **2016**, *63*, 62–92, DOI: 10.1016/j.biortech.2010.09.093.

(5) Juan, J. C.; Kartika, D. A.; Wu, T. Y.; Hin, T.-Y. Y. Biodiesel production from jatropha oil by catalytic and non-catalytic approaches: an overview. *Bioresource Technology* **2011**, *102*, 452–460, DOI: 10.1016/j.rser.2016.05.035.

(6) Aransiola, E.; Ojumu, T.; Oyekola, O.; Madzimbamuto, T.; Ikhu-Omoregbe, D. A review of current technology for biodiesel production: State of the art. *Biomass and Bioenergy* **2014**, *61*, 276–297, DOI: 10.1016/j.biombioe.2013.11.014.

(7) Leung, D. Y.; Wu, X.; Leung, M. A review on biodiesel production using catalyzed transesterification. *Applied Energy* **2010**, *87*, 1083–1095, DOI: 10.1016/j.apenergy.2009.10.006.

(8) Sadeghinezhad, E.; Kazi, S.; Badarudin, A.; Oon, C.; Zubir, M.; Mehrali, M. A comprehensive review of bio-diesel as alternative fuel for compression igni-

tion engines. *Renewable and Sustainable Energy Reviews* **2013**, *28*, 410–424, DOI: 10.1016/j.rser.2013.08.003.

(9) Mercuri, E.; Manca, D.; Abderazag, O.; Patel, R.; Mujtaba, I. Biodiesel from Sunflower Oil: Development of Process Model via Lab scale Experiment and Optimisation. *Computer Aided Chemical Engineering* **2016**, *38*, 1713–1718, DOI: 10.1016/B978-0-444-63428-3.50290-3.

(10) Abbasi, S.; Diwekar, U. M. Characterization and stochastic modeling of uncertainties in the biodiesel production. *Clean Technologies and Environmental Policy* **2014**, *16*, 79–94, DOI: 10.1007/s10098-013-0596-4.

(11) Benavides, P. T.; Salazar, J.; Diwekar, U. Economic comparison of continuous and batch production of biodiesel using soybean oil. *Environmental Progress & Sustainable Energy* **2013**, *32*, 11–24, DOI: 10.1002/ep.11736.

(12) Zhang, Y.; Dube, M.; McLean, D.; Kates, M. Biodiesel production from waste cooking oil: 1. Process design and technological assessment. *Bioresource Technology* **2003**, *89*, 1–16, DOI: 10.1016/S0960-8524(03)00040-3.

(13) Philip, F. A.; Veawab, A.; Aroonwilas, A. Simulation Analysis of Energy Performance of Distillation-, Stripping-, and Flash-Based Methanol Recovery Units for Biodiesel Production. *Industrial & Engineering Chemistry Research* **2014**, *53*, 12770–12782, DOI: 10.1021/ie5003476.

(14) Jurado, M.; Plesu, V.; Ruiz, J.; Ruiz, A.; Tuluc, A.; Llacuna, J. Simulation of a hybrid reactive extraction unit: biodiesel synthesis. *Chemical Engineering Transactions* **2013**, *33*, 205–210, DOI: 10.3303/CET1335034.

(15) GutieÌĄrrez-Arriaga, C. G.; Serna-González, M.; Ponce-Ortega, J. M.; El-Halwagi, M. M. Sustainable integration of algal biodiesel production with steam electric

power plants for greenhouse gas mitigation. *ACS Sustainable Chemistry & Engineering* **2014**, *2*, 1388–1403, DOI: 10.1021/sc400436a.

(16) Martín, M.; Grossmann, I. E. ASI: Toward the optimal integrated production of biodiesel with internal recycling of methanol produced from glycerol. *Environmental Progress & Sustainable Energy* **2013**, *32*, 891–901, DOI: 10.1002/ep.11836.

(17) Martín, M.; Grossmann, I. E. Simultaneous optimization and heat integration for biodiesel production from cooking oil and algae. *Industrial & Engineering Chemistry Research* **2012**, *51*, 7998–8014, DOI: 10.1021/ie2024596.

(18) Moradi-Aliabadi, M.; Huang, Y. Multistage optimization for chemical process sustainability enhancement under uncertainty. *ACS Sustainable Chemistry & Engineering* **2016**, *4*, 6133–6143, DOI: 10.1021/acssuschemeng.6b01601.

(19) Patle, D. S.; Sharma, S.; Ahmad, Z.; Rangaiah, G. Multi-objective optimization of two alkali catalyzed processes for biodiesel from waste cooking oil. *Energy Conversion and Management* **2014**, *85*, 361–372, DOI: 10.1016/j.enconman.2014.05.034.

(20) Fahmi, I.; Cremaschi, S. Process synthesis of biodiesel production plant using artificial neural networks as the surrogate models. *Computers & Chemical Engineering* **2012**, *46*, 105–123, DOI: 10.1016/j.compchemeng.2012.06.006.

(21) Komers, K.; Skopal, F.; Stloukal, R.; Machek, J. Kinetics and mechanism of the KOHâĂŤcatalyzed methanolysis of rapeseed oil for biodiesel production. *European Journal of Lipid Science and Technology* **2002**, *104*, 728–737, DOI: 10.1002/1438-9312(200211)104:11<728::AID-EJLT728>3.0.CO;2-J.

(22) Vicente, G.; Martinez, M.; Aracil, J. Kinetics of Brassica c arinata Oil Methanolysis. *Energy & Fuels* **2006**, *20*, 1722–1726, DOI: 10.1021/ef060047r.

(23) Bambase, M. E.; Nakamura, N.; Tanaka, J.; Matsumura, M. Kinetics of hydroxide-catalyzed methanolysis of crude sunflower oil for the production of fuel-grade methyl esters. *Journal of Chemical Technology and Biotechnology* **2007**, *82*, 273–280, DOI: 10.1002/jctb.1666.

(24) Bikou, E.; Louloudi, A.; Papayannakos, N. The effect of water on the transesterification kinetics of cotton seed oil with ethanol. *Chemical Engineering & Technology* **1999**, *22*, 70–75, DOI: 10.1002/(SICI)1521-4125(199901)22:1<70::AID-CEAT70>3.0.CO;2-0.

(25) Marjanović, A. V.; Stamenković, O. S.; Todorović, Z. B.; Lazić, M. L.; Veljković, V. B. Kinetics of the base-catalyzed sunflower oil ethanolysis. *Fuel* **2010**, *89*, 665–671, DOI: 10.1016/j.fuel.2009.09.025.

(26) Richard, R.; Thiebaud-Roux, S.; Prat, L. Modelling the kinetics of transesterification reaction of sunflower oil with ethanol in microreactors. *Chemical Engineering Science* **2013**, *87*, 258–269, DOI: 10.1016/j.ces.2012.10.014.

(27) Noureddini, H.; Zhu, D. Kinetics of transesterification of soybean oil. *Journal of the American Oil Chemists' Society* **1997**, *74*, 1457–1463, DOI: 10.1007/s11746-997-0254-2.

(28) Bashiri, H.; Pourbeiram, N. Biodiesel production through transesterification of soybean oil: A kinetic Monte Carlo study. *Journal of Molecular Liquids* **2016**, *223*, 10–15, DOI: 10.1016/j.molliq.2016.08.029.

(29) Stamenković, O. S.; Todorović, Z. B.; Lazić, M. L.; Veljković, V. B.; Skala, D. U. Kinetics of sunflower oil methanolysis at low temperatures. *Bioresource Technology* **2008**, *99*, 1131–1140, DOI: 10.1016/j.biortech.2007.02.028.

(30) Vicente, G.; Martínez, M.; Aracil, J.; Esteban, A. Kinetics of sunflower oil methanol-

ysis. *Industrial & Engineering Chemistry Research* **2005**, *44*, 5447–5454, DOI: 10.1021/ie040208j.

(31) Darnoko, D.; Cheryan, M. Kinetics of palm oil transesterification in a batch reactor. *Journal of the American Oil Chemists' Society* **2000**, *77*, 1263–1267, DOI: 10.1007/s11746-000-0198-y.

(32) Shahbazi, M. R.; Khoshandam, B.; Nasiri, M.; Ghazvini, M. Biodiesel production via alkali-catalyzed transesterification of Malaysian RBD palm oil–characterization, kinetics model. *Journal of the Taiwan Institute of Chemical Engineers* **2012**, *43*, 504–510, DOI: 10.1016/j.jtice.2012.01.009.

(33) Issariyakul, T.; Dalai, A. K. Comparative kinetics of transesterification for biodiesel production from palm oil and mustard oil. *The Canadian Journal of Chemical Engineering* **2012**, *90*, 342–350, DOI: 10.1002/cjce.20679.

(34) Likozar, B.; Levec, J. Effect of process conditions on equilibrium, reaction kinetics and mass transfer for triglyceride transesterification to biodiesel: experimental and modeling based on fatty acid composition. *Fuel Processing Technology* **2014**, *122*, 30–41, DOI: 10.1016/j.fuproc.2014.01.017.

(35) Likozar, B.; Levec, J. Transesterification of canola, palm, peanut, soybean and sunflower oil with methanol, ethanol, isopropanol, butanol and tert-butanol to biodiesel: Modelling of chemical equilibrium, reaction kinetics and mass transfer based on fatty acid composition. *Applied Energy* **2014**, *123*, 108–120, DOI: 10.1016/j.apenergy.2014.02.046.

(36) Eze, V. C.; Phan, A. N.; Harvey, A. P. A more robust model of the biodiesel reaction, allowing identification of process conditions for significantly enhanced rate and water tolerance. *Bioresource Technology* **2014**, *156*, 222–231, DOI: 10.1016/j.biortech.2014.01.028.

(37) Encinar, J.; Gonzalez, J.; Rodriguez, J.; Tejedor, A. Biodiesel fuels from vegetable oils: transesterification of Cynara c ardunculus L. oils with ethanol. *Energy & fuels* **2002**, *16*, 443–450, DOI: 10.1021/ef010174h.

(38) Berchmans, H. J.; Morishita, K.; Takarada, T. Kinetic study of hydroxide-catalyzed methanolysis of Jatropha curcas–waste food oil mixture for biodiesel production. *Fuel* **2013**, *104*, 46–52, DOI: 10.1016/j.fuel.2010.01.017.

(39) Tiwari, P.; Kumar, R.; Garg, S. Transesterification, modeling and simulation of batch kinetics of non-edible vegetable oils for biodiesel production. *Annu Alche Meet, San Francisco* **2006**, DOI: N/A.

(40) Rashid, U.; Anwar, F. Production of biodiesel through optimized alkaline-catalyzed transesterification of rapeseed oil. *Fuel* **2008**, *87*, 265–273, DOI: 10.1016/j.fuel.2007.05.003.

(41) Costa-Felix, R. P.; Ferreira, J. R. Comparing ultrasound and mechanical steering in a biodiesel production process. *Physics Procedia* **2015**, *70*, 1066–1069, DOI: 10.1016/j.phpro.2015.08.227.

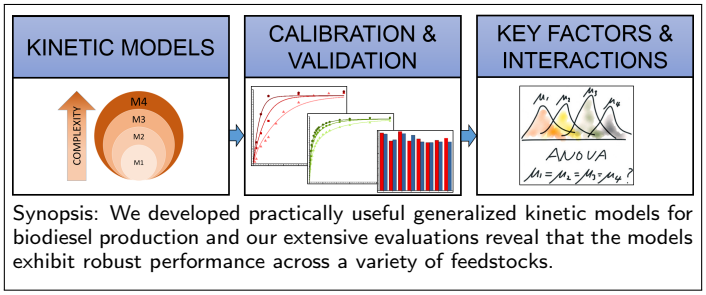
(42) Reyero, I.; Arzamendi, G.; Zabala, S.; Gandía, L. M. Kinetics of the NaOH-catalyzed transesterification of sunflower oil with ethanol to produce biodiesel. *Fuel Processing Technology* **2015**, *129*, 147–155, DOI: 10.1016/j.fuproc.2014.09.008.

(43) Andrikopoulos, N. K. Triglyceride species compositions of common edible vegetable oils and methods used for their identification and quantification. *Food Reviews International* **2002**, *18*, 71–102, DOI: 10.1081/FRI-120003418.

(44) Garud, S. S.; Karimi, I. A.; Kraft, M. Smart sampling algorithm for surrogate model development. *Computers & Chemical Engineering* **2017**, *96*, 103–114, DOI: 10.1016/j.compchemeng.2016.10.006.

(45) Pinzi, S.; Mata-Granados, J.; Lopez-Gimenez, F.; de Castro, M. L.; Dorado, M. Influence of vegetable oils fatty-acid composition on biodiesel optimization. *Bioresource Technology* **2011**, *102*, 1059–1065, DOI: 10.1016/j.biortech.2010.08.050.

Graphical TOC Entry



Synopsis: We developed practically useful generalized kinetic models for biodiesel production and our extensive evaluations reveal that the models exhibit robust performance across a variety of feedstocks.